12-1-2013

Differences between conventional and glyphosate tolerant soybeans and moisture effect in their discrimination by near infrared spectroscopy

Lidia Esteve Agelet
Iowa State University, esteve.lidia@gmail.com

Paul R. Armstrong
United States Department of Agriculture

Jasper G. Tallada
United States Department of Agriculture

Charles R. Hurburgh Jr.
Iowa State University, tatry@iastate.edu

Differences between conventional and glyphosate tolerant soybeans and moisture effect in their discrimination by near infrared spectroscopy

Abstract

Previous studies showed that Near Infrared Spectroscopy (NIRS) could distinguish between Roundup Ready® (RR) and conventional soybeans at the bulk and single seed sample level, but it was not clear which compounds drove the classification. In this research the varieties used did not show significant differences in major compounds between RR and conventional beans, but moisture content had a big impact on classification accuracies. Four of the five RR samples had slightly higher moistures and had a higher water uptake than their conventional counterparts. This could be linked with differences in their hulls, being either compositional or morphological. Because water absorption occurs in the same region as main compounds in hulls (mainly carbohydrates) and water causes physical changes from swelling, variations in moisture cause a complex interaction resulting in a large impact on discrimination accuracies.

Keywords

NIR, Soybeans, Roundup ready, Genetically modified organisms

Disciplines

Agriculture | Bioresource and Agricultural Engineering

Comments

This article is from *Food Chemistry* 141 (2013): 1895–1901, doi:10.1016/j.foodchem.2013.04.087.

Food Chemistry 141 (2013) 1895–1901

Contents lists available at SciVerse ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Differences between conventional and glyphosate tolerant soybeans and moisture effect in their discrimination by near infrared spectroscopy ☆

Lidia Esteve Agelet [a,*], Paul R. Armstrong [b], Jasper G. Tallada [b], Charles R. Hurburgh Jr. [c]

[a] Department of Agricultural and Biosystems Engineering, Iowa State University, United States
[b] Engineering and Wind Erosion Research Unit, USDA-ARS, Center for Grain and Animal Health Research, 1515 College Ave, Manhattan, KS, United States
[c] Department of Agriculture and Biosystems Engineering, Iowa State University, United States

ARTICLE INFO

Article history:
Received 18 February 2013
Received in revised form 20 April 2013
Accepted 25 April 2013
Available online 7 May 2013

Keywords:
NIR
Soybeans
Roundup ready
Genetically modified organisms

ABSTRACT

Previous studies showed that Near Infrared Spectroscopy (NIRS) could distinguish between Roundup Ready® (RR) and conventional soybeans at the bulk and single seed sample level, but it was not clear which compounds drove the classification. In this research the varieties used did not show significant differences in major compounds between RR and conventional beans, but moisture content had a big impact on classification accuracies. Four of the five RR samples had slightly higher moistures and had a higher water uptake than their conventional counterparts. This could be linked with differences in their hulls, being either compositional or morphological. Because water absorption occurs in the same region as main compounds in hulls (mainly carbohydrates) and water causes physical changes from swelling, variations in moisture cause a complex interaction resulting in a large impact on discrimination accuracies.

1. Introduction

Genetically modified (GM) organisms have been manipulated to avoid diseases, to enhance resistance to herbicides, and to increase their nutritional value. Not all world markets fully accept GM products however, for a variety of reasons regarding the introduction of new allergens, possible development of antibiotic-resistant bacteria strains and environmental biodiversity issues (Cohen, Chang, Boyer, & Helling, 1973; Bakshi, 2003). Many countries have set regulations for identification, quantification, and appropriate labeling of products containing GM organisms. Roundup® is a popular glyphosate-based herbicide intended to kill a broad variety of plants on contact. Roundup® application in crops used to be only possible at certain developmental stages without direct application (Benbrook, 2009). The development of herbicide resistant crops reduced these restrictions. The patenting and marketing of Roundup® resistant crops, licensed with the name of Roundup Ready®, was initially done by Monsanto in 1996 (Patent EP 546090). Using genetic recombinant DNA technology, genetic material from the bacteria *Agrobacterium tumefacien* was introduced to the crop genoma, conferring the crop a high tolerance to the herbicide. This removed many restrictions on Roundup® use, lowered production costs and increased crop yields (Schenepf, 2003).

Soybeans (*Glycine max* L.) were the first Roundup Ready® (RR) crop to be introduced into markets in 1996. They rapidly displaced conventional soybeans because of advantages for crop management and yields, and currently account for more than half of the soybean fieldcrops around the world (Konduru, Kruse, & Kalaitzandonakes, 2008). RR soybeans are widely accepted in the global markets; they are one of the two currently accepted GM varieties of soybeans in Europe, which has the most restrictive laws regarding GM importation. But despite of their acceptance in many markets, they must be labeled as a GM crop, even if they are present as adventitious contamination in conventional batches whenever their percentage exceeds pre-established thresholds. Current thresholds of adventitious GM contamination in conventional soybeans for feeding purposes range from 0.9% for Europe to 5% for Japan and Taiwan. In the case of Europe, the tolerance limit applies to contamination of recognised GM events; otherwise, the threshold is reduced to 0.5% if the events are proven safe, even if not politically accepted.

Determination of GM contamination levels for large shipments is challenging. Current methods are time consuming, complex, and not suitable for rapid on-site measurements because of laboratory-based analysis. The analyses are divided into protein-based methods and DNA-based methods. Protein-based methods such as Enzyme-Linked Immuno Sorbent Assay (ELISA) use specific antibodies, which require previous knowledge of the GM to be

☆ Mention of trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

* Corresponding author. Address: 1545 Food Science Building, Iowa State University, Ames, IA 50014, United States. Tel.: +1 515 294 8319.

E-mail address: esteve.lidia@gmail.com (L. Esteve Agelet).

0308-8146/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2013.04.087

analysed, but are quicker, cheaper, and simpler than DNA-based methods (Konduru et al., 2008). DNA-based methods are laboratory sensitive, expensive and slow; the lowest practical limit of detection of GM DNA material is around 0.1% (Miljuš-Djukić et al., 2010). These methods are destructive meaning that even in the case they could perform faster in an on-line setting, only a small portion of the sample could be analysed. This leads to the added problem of taking representative samples from large shipments, proven to be a function of grain type and the threshold to be analysed (Hübner, Waiblinger, Pietsch, & Brodmann, 2001).

Near infrared spectroscopy cannot be used to analyse trace elements or genetic information, but it can measure changes in structure or concentration of organic compounds that can be the fruit of the phenotypic expression of genes. Several researches reported that no significant differences in concentration of major biochemical compounds (protein, oil, or fiber among others) exist between conventional and their corresponding transgenic counterparts (Harrigan et al., 2010; McCann, Liu, Trujillo, & Dobert, 2005; Taylor, Fuchs, MacDonald, Shariff, & Padgette, 1999) . However, some researchers suggest differences in minor compounds such as length of chain acids (Jimenez, Bernal, Nozal, Toribio, & Bernal, 2009) or other unintended (pleiotrofic) effects from the genetically modification of RR which may be noticeable in specific varieties or specific environmental conditions. For instance, RR crops were found to suffer higher weight loss under water shortage (Gertz, Vencill, & Hill, 1999). Any of those side-effects from introducing the gene of RR resistance may have lead Roussel et al. to use Near Infrared Spectroscopy (NIRS) to discriminate RR and conventional soybeans by NIR transmittance with notable success (Roussel, Hardy, Hurburgh, & Rippke, 2001). Over 3000 bulk soybean samples were scanned from each class (RR and conventional) by transmittance instruments. Non-linear classification methods such as locally weighted principal component regression (LW-PCR) and artificial neural network (ANN) were used to achieve classification accuracies of 93% and 88% respectively, with validation and training sets from a single crop-year and combining two spectrophotometers. Two recent studies were carried out to analyse the discrimination by diffuse reflectance NIRS at single seed level. Lee and Choung (2011) carried out a feasibility study involving 10 samples of conventional soybeans (50 seeds per sample) and 10 herbicide resistant soybeans (50 seeds per sample). Accuracies of 97% were achieved utilising NIR and visible (VIS) radiation with Partial Least Squares Discriminant Analysis (PLS-DA), although major spectral differences between the two classes in the study arise in the VIS region. A more recent study by Agelet, Gowen, Hurburgh, and O'Donell (2012) was conducted involving three reflectance NIR technologies, over 240 samples from several crop-years (over 3000 beans total), and two validation sets: (1) new seeds from samples represented in the training set, and (2) a validation set with seeds from new samples, not included in the training set. Algorithms used for discrimination were those from the first study with bulk samples and transmittance NIR: LW-PCR and ANN (Roussel et al., 2001). Best discrimination accuracies were achieved with LW-PCR when beans for validation belonged to samples represented in the training set (up to 94%), similar to Lee and Choung (2011) results. Results were worse with ANN (lower 80% range). Those results were very similar to those of Roussel et al. for bulk samples (Roussel et al., 2001). When new beans in the validation set were from samples not represented in the training set, classification accuracies dropped to the mid 70% on average. However, a few questions remain about what is detected by NIRS in differentiating RR and conventional varieties. Both Roussel et al. (2001) and Agelet, Gowen, Hurburgh, and O'Donell (2012) suggested the carbohydrate region to be the most influential in the discrimination of large sets of samples. Lee and Choung (2011), on the other hand, found differences in color and several regions related to CO_2H, CCl, and H_2O. Roussel et al. (2001)) also mentioned that both RR and conventional misclassified samples in their research had moisture content higher than 13%.

In this research we took a closer look at the differences between 5 conventional soybean varieties and their RR counterparts, regarding their chemical composition and what is detected by NIRS. We utilised two NIR instruments for the latest task: A Fourier-Transform Near Infrared Transmittance (FT-NIR) (NIRFlex N-500 by Buchi Corporation) and the USDA light-tube, working by diffuse reflectance (Armstrong, 2006; Tallada, Palacios-Rojas, & Armstrong, 2009). We proceeded to study the impact of moisture changes in the discrimination accuracies. Since beans destined to elevators and commodities have variable moisture, any effect that moisture may induce to the discriminative ability of the models should be taken in consideration.

2. Materials and methods

2.1. Samples

Five conventional public soybean varieties from 2007 crop-year (labeled as M97-302, M97-303, M97-304, M97-305, and M97-306 varieties) and their same respective varieties with the Roundup Ready® (RR) gene were used in this study. This made a set of 10 samples. The samples were harvested from a same location (Iowa State University Curtiss farm, Ames, IA), and neither the plant nor the seeds received chemical treatments. Samples belonged to the same crop year in order to reduce phenotype–environment interactions.

One hundred and fifty seeds were selected from each sample and scanned by the two instruments consecutively (1500 scanned seeds total). The initial average moisture of the bulk seeds was measured by with an Infratec 1221 transmittance instrument (Foss North America, Eden Prairie, MN, USA) using a cuvette and the Iowa State moisture calibration. Sample composition predicted with ANN Iowa State calibrations as shown in Table 1. The standard error of prediction (SEP) in an independent validation of the moisture calibration was 0.37%, SEP = 0.52% for protein, SEP = 0.37% for oil, and SEP = 0.08% for fiber.

Additional sets of 150 seeds from each sample were selected and sealed in individual small plastic bags with a wet paper towel on the top, avoiding direct contact with the seeds to avoid spoilage. The paper towels were cut from conventional disposable laboratory cellulose towels of one sheet, measuring 3×3 cm of surface. The sealed bags were kept at 2 °C for around 3 weeks or until their average moisture was over 13%. The moisture on the seeds was monitored and predicted with the Infratec 1221 instrument. During that period of time, the paper towels were replaced when were slightly dump at touch, and seeds were shaken to allow better equilibration of moisture within the samples. After scanning each

Table 1
Bulk composition of the 10 samples used in the study predicted with NIR transmittance and Iowa State calibrations.

Sample	Initial Moisture (%)	Protein[a] (%)	Oil[a] (%)	Fiber[a] (%)
M97-302 RR	8.8	34.9	16.9	5.0
M97-302	8.6	36.1	18.1	4.8
M97-303 RR	8.4	36.0	18.8	4.7
M97-303	8.4	36.4	17.4	4.8
M97-304 RR	8.2	37.9	17.0	4.7
M97-304	8.3	36.2	18.0	4.8
M97-305 RR	9.3	38.0	17.3	4.6
M97-305	8.9	36.3	17.9	4.8
M97-306 RR	9.5	36.2	18.3	4.7
M97-306	8.9	34.6	18.0	4.7

[a] 13% Moisture content basis.

seed in the two single seed instruments consecutively, the moisture of each seed was determined by oven drying for 3 h at 130 °C (AOCS, Ac 2-41 method).

2.2. Instrumentation

Two spectrometers were used. Buchi NIRFlex N-500 (Buchi Corporation, New Castle, DE) is a Near Infrared Fourier Transform (FT-NIR) spectrometer which has the capability of working in both reflectance and transmittance mode with the use of appropriate modules. For this research, the NIRFlex solids transmission module with the 10-well sample cell was used to analyse individual seeds. The instrument covers the spectral range from 11,520 to 6000 cm^{-1} (868.1–1666.7 nm), with a 4 cm^{-1} sampling increment (1381 data points) at full resolution of 8 cm^{-1}. The second spectrometer was designed and built by the United States Department of Agriculture (USDA) at the Center for Grain and Animal Health Research, Manhattan (KS). This unit differs from conventional single point spectrometers in the fact that it attempts to collect spectra from the entire seed surface. The light-tube instrument sample cell is a borosilicate glass tube in which a single seed falls through while being illuminated by 48 miniature tungsten lamps located around the periphery of the tube. A photoelectric switch D12DAB6FP (Banner Engineering Corp., Minneapolis, MN) detects when the seed is manually dropped through the tube from a small funnel located on the top of the tube which initiates spectral measurement. Spectral reflectance from the whole seed is collected through a Y-shaped bifurcated fiber optic BIF600-VIS–NIR (Ocean Optics, Dunedin Fla.) with two of the ends attached at each end of the tube. The receiving end of the fiber optic is connected to the CDI spectrometer model NIR256-1.7T1-USB2/3.1/50 μm SNIR 1074 (Control Development, Inc., South Bend, Ind.) at 1 nm sampling increments from 904 to 1686 nm. More information related to the instrument can be found in the literature (Armstrong, 2006; Tallada, Palacios-Rojas, & Armstrong, 2009). A background and reference measurement was taken every 20 min with the reference measurement using an empty illuminated tube. Each seed spectrum was the average of three spectra after mean-centering each individual spectrum.

2.3. Data management and discrimination models

Wavelengths from the two extremes ends of the spectra were removed for the light-tube instrument to remove the regions with lower detector sensitivity. The working spectra from the light-tube covered the range from 953 to 1636 nm (50 nm removed from each end on each single seed spectrum). The working range from the FT-NIR was reduced to 868.1–1100 nm, which is the optimal wavelength range for transmittance measurements in grains. Detection and outlier removal was done by visual inspection of the spectra, and principal component analysis (PCA) within varieties. Samples showing either extreme scores in the first 18 principal components (PCs) or high leverage vs. high residual variance values were flagged and removed as possible outliers. No preprocessing of the spectra was done based on previous studies that showed robustness and smoothing methods also remove relevant information used for conventional and RR discrimination (Roussel et al., 2001). Data management and discrimination models were developed in Matlab 7.10.0 (2010a) (Mathworks, Natick, MA) and PLS_toolbox v.5.8.3 (Eigenvector Research Inc., Wenatchee, WA).

The SIMPLS algorithm for Partial Least Squares Discriminate Analysis (PLS-DA) was used for all analyses, since it worked well for our study when only a few varieties are involved. This was similarly shown by Lee and Choung (2011). The algorithm compresses spectral data to a few uncorrelated variables (latent variables) which are calculated following both the direction of maximum variability of the data while keeping the maximum information related to the seed classes. The dependent variable or seed class was designated 1 and 2, representing conventional and RR beans, respectively. The threshold for discriminating classes was determined by the PLS-DA function through Bayesian inference which assumes a normal distribution of prediction for each class. Using leave-one-out cross-validation, the optimal number of latent variables for the final model was selected by looking at the fractional misclassification rate of each class and the minimum total misclassification rate.

2.4. Experimental design

Differences between and within varieties. Varietal differences were analysed from a compositional point of view and by NIR means. Three attributes of the beans were determined by NIR single seed calibrations in the FT-NIR transmittance spectrometer: seed mass (mg), oil content (% dry weight), and protein content (% as is). The models had been previously validated with independent validation sets, leading to the following statistics: Standard Error of Prediction (SEP) = 0.64%, dry weight basis, and RPD = 3.34 for oil; SEP = 1.00%, as-is weight basis, and RPD = 2.02 for protein; SEP = 13.7 mg and RPD = 3.18 for seed weight.

Three two-way ANOVA analyses for each attribute (protein, oil, and weight) were performed using the following model:

$$Y = \alpha_i + \beta_i + \alpha_i \cdot \beta_i + e_i$$

where Y is the dependent variable (weight, oil, or protein), α is the genetic trait independent variable (conventional or RR), β is the independent variable corresponding to the variety (5 varieties in our study), $\alpha_i \cdot \beta_i$ is the interaction term between genetic trait and variety, and e is the error term. The variety factor was considered as a random effect when entering the model into the statistical program IBM SPSS Statistics 20 software (Armonk, NY, USA). Results of the ANOVA helped to determine if there were significant overall differences in weight, oil content, and protein content between conventional and RR classes, and differences between RR and conventional seeds within varieties. For the last test, we used *T*-tests to compare means and the *F*-test to determine the equality of variances between classes for each variety, using Microsoft Excel 2007 statistical functions.

Five NIR PLS-DA models were developed to discriminate conventional and RR soybeans for each variety. In each single-variety model, two-thirds of the data was used for training (approx. 200 spectra from 200 beans) and one-third for validation (approx. 100 spectra from 100 beans). Each model from each variety was tested with the test set from the other varieties.

2.5. Testing the moisture effect

A discrimination model was created for each instrument with all the beans from varieties M97-302, M97-303, M97-305, and M97-306, leading to a training set of around 600 spectra. Variety M97-304 was left out as a test set (150 spectra), since it was seen to differ most from the others within the variety discrimination exercise.

These general models were then used to predict the new beans with higher moisture. The initial plan was to develop classification models based on moisture ranges and an overall model with a wide moisture range. However, the final moisture was not homogeneous between and within samples as we later report. For this reason we carried out two small tests with fewer beans. The first validation with higher moisture seeds was performed using seeds from a single variety (M97-304, later shown to be the variety with highest misclassifications). The M97-304 tested seeds belonged to different moisture ranges (range 1: seven beans with up to 10%; range

Table 2
Summary of single seed prediction statistics of oil, protein, and weight for each sample. Shadowed cells indicate the average or variance (SD cells) conventional-RR pairs per variety that were statistically different at 5% significance level.

Sample	Protein (%)		Oil (%)		Weight (mg)	
	Average	SD	Average	SD	Average	SD
M97-302 RR	38.01	2.24	18.47	1.32	120.7	23.1
M97-302	37.56	3.56	19.45	1.97	116.0	21.1
M97-303 RR	37.78	2.64	19.65	1.48	113.8	25.3
M97-303	38.41	2.56	19.05	1.39	114.4	24.0
M97-304 RR	39.87	2.53	18.78	1.80	122.0	24.4
M97-304	38.34	2.80	19.57	1.84	137.2	25.7
M97-305 RR	39.85	2.23	18.04	1.28	130.1	24.7
M97-305	39.37	3.00	18.87	1.46	126.5	23.1
M97-306 RR	38.02	2.54	19.38	1.20	139.1	22.1
M97-306	37.38	2.54	19.48	1.49	122.5	19.5

2: 28 beans with moisture from 10.1% to 13.5%; range 3: 44 beans with moisture from 13.6% to 15%) in order to test how small changes in moisture may affect the classification results. Summarising, the intent was to determine if there was any correlation between moisture content of the seeds and classification accuracy. A second validation set included seeds from all the varieties (40 seeds per variety, 20 conventional and 20 RR), covering a wider high moisture range of 13.5–17%. The number of seeds from each test set was again limited by the heterogeneous moisture observed for each individual sample during the process of increasing sample moisture. With this test we wanted to determine if a moisture effect would impact all the varieties the same way.

It seems to be reasonable to assume as a first hypothesis, and in absence of any interaction with moisture content, that since the moisture range from the last two test sets was not covered by the discrimination models then the discrimination accuracies should be low and would equally impact RR and conventional classes. That is to say, the number of misclassified beans from RR and conventional classes should not be significantly different.

3. Results and discussion

3.1. Compositional differences between conventional and Roundup Ready samples

Table 2 shows the prediction descriptive statistics for each variety and class (RR and conventional). The single seed calibrations for oil and protein may have slight biases of around one percentage point as this is the average difference with the bulk sample predictions.

3.1.1. Variance

F-tests were conducted for each sample and predicted compound in order to compare significant differences in variances between RR and conventional within each variety at a significance level of 0.01%. The pairs that were found to differ in their variances have their standard deviations (SD) highlighted (dark) in Table 2. Seven of the pairs had significant differences in their variance. This impacted the ANOVA analysis, as the assumption of homoscedasticity in ANOVA was violated according to the rejection of the hypothesis of the same variance between groups (Levene's test) at 0.01 significance level. No data transformation could lead to similar variances based on Levene's test. However, since the group sample size is large (>600) and very close between groups, the ANOVA is more robust to violations of normality and homocedasticity assumptions while the analysis has more power (Underwood, 1997).

3.1.2. Average

For all three attributes, the effect of the genetic trait (RR or conventional) was not significant at 0.05 significant level. The overall average for oil was 18.87% for RR and 19.03% for conventional beans. For protein, the average was 38.68% for RR and 38.15% for conventional beans. The average weight for RR beans was 125.1 mgs and 123.3 mgs for conventional beans. Therefore, there are not overall significant differences between means for RR and conventional samples for protein, oil, or weight. For weight, a previous study (Elmore et al., 2001) indicated that conventional seeds weigh more, but there was no significant difference for the varieties involved in our study with the average weight of conventional slightly lower than RR beans. The factor variety was not significant either, meaning that all varieties involved in the study have similar concentration of oil and protein, and similar weight. The interaction between the genetic trait (RR and conventional) and variety was significant for all three attributes, which implies that in some varieties there are differences in oil, protein, or weight between the conventional and RR seeds. The averages within varieties which are significantly different according a two-tailed *t*-test are shadowed (light) in Table 2. The assumption of unequal variance has been considered following the results of the *F*-test for equality of variances (significantly different variances are also shadowed in Table 2).

From the bulk sample predictions in Table 1, we could expect no differences in fiber content or moisture between RR and conventional samples either, given the fact that the variability and range of any of the previously analysed compounds (oil, protein, weight) within a sample is usually large (Armstrong, 2006), and yet no significant difference between RR and conventional classes was found in this study. These findings agree with other research indicating that there are no major differences in main composition between conventional and RR soybeans (Harrigan et al., 2010; McCann et al., 2005; Taylor et al., 1999). Differences in single-seed moisture between RR and conventional beans for samples M97-305 and M97-306 may be significant, with RR samples having higher moisture content, but since no individual moisture was taken for each bean it could not be tested.

3.2. Differences between conventional and Roundup Ready samples by NIRS

Previous research has shown mixed results regarding the absorbance of RR and conventional soybeans. while studies from Agelet, et al. (2012) and Roussel et al. (2001) found that the average spectra of RR seeds had higher overall apparent absorbance than conventional, Lee and Choung (2011) observed the opposite. In this current research, only variety M97-305 showed higher absorbance average values for RR seeds for the light tube. The opposite was observed for the FT transmittance instrument: all varieties showed the RR seeds with higher absorbance compared to conventional

Table 3
Table of misclassified beans using single-variety PLS-DA models from both FT-NIR and light tube to discriminate beans from the five varieties (columns). C refers to conventional misclassified beans, and RR to Roundup Ready beans. The first number is the misclassified beans from the FT-NIR model, and the number after the slash is the misclassified beans from the light-tube model.

Model/ test data	M97-302 set	M97-303 set	M97-304 set	M97-305 set	M97-306 set
M97-302 model	C:0/0 RR:0/0	C:15/35 RR:15/45	C:0/40 RR:50/0	C:0/50 RR:0/48	C:15/50 RR:0/0
M97-303 model	C:0/50 RR:0/10	C:4/0 RR:3/0	C:0/35 R:50/15	C:0/4 RR:0/20	C:3/30 RR:0/50
M97-304 model	C:50/1 RR:0/45	C:50/30 RR:0/25	C:2/4 RR:1/0	C:50/50 RR:0/0	C:50/48 RR:0/1
M97-305 model	C:0/25 RR:50/50	C:0/35 RR:50/49	C:0/2 RR:50/50	C:0/0 RR:0/0	C:0/1 RR:13/0
M97-306 model	C:0/0 RR:30/50	C:0/35 RR:50/48	C:0/0 RR:50/50	C:0/0 RR:0/20	C:0/0 RR:0/0

seeds. This may be due to the instrument inner characteristics. We conclude that either higher or lower absorbance overall values cannot be linked to the phenotypical expression of RR genes in soybean as it depends on the varieties under analysis and the instrument involved.

The discrimination by PLS-DA for each variety was possible with high accuracies overall. The number of latent variables needed for all models from both instruments ranged from 5 to 7. Table 3 shows the number of misclassified beans for conventional (C) and RR seeds for discrimination models developed for each variety (rows) and validated with the test sets (columns). The test sets included beans from the same variety involved in the training set and also beans from the remaining 4 varieties. For both C and RR misclassifications in the table, the first number are misclassified beans from FT-NIR models, the following number after the slash is light tube model misclassifications. Both instruments performed similar when discriminating beans from the varieties involved in the test set (i.e., column 1, row 1; column 2, row 2;). Almost all validation beans were correctly classified by the models developed from beans within the same variety (Table 3). This indicates that each class, even if there is overlap of beans with the same content of the major compounds, is easily differentiated by NIRS as two different samples. Breeders could benefit of the NIRS technology for segregating RR and conventional beans from a same variety using principal component analysis with unsupervised classification methods because the two classes show a good separation on some of the latent variable scores (figures not shown).

Varieties M97-303 and M97-304 had the most misclassified beans (3 beans from the FT-NIR instrument and 4 from the light tube) even if according to the single bean predictions there were significant differences in one or more of the 3 measured attributes (weight, oil, and protein content) between RR and conventional classes (Table 2). On the other hand, varieties such as M97-306 which only were found to have a significant difference in seed weight between conventional and RR seeds, could be all successfully discriminated.

Varieties M97-305 and M97-306 behave similarly, as models developed with one of the two varieties could be used to perfectly discriminate beans from the other (Table 3). Their protein and oil content, however, were not the closest among the rest of varieties (Tables 1 and 2). Nevertheless, moisture content for both conventional and RR beans in both varieties had similar values, with RR having slightly higher moisture content than conventional beans. M97-302 variety also followed that trend, and could also be discriminated with success by the M97-305 and M97-306 models in the FT-NIR instrument. The trend was not exactly the same for the light tube instrument. Models developed from the light tube validated with beans from another variety lead to more random classification results and often more misclassified seeds. One possible explanation is that the light tube models are more specific to the seeds utilised for training, not only based on their composition but also on their physical characteristics. The suspicious of this phenomenon was also commented on in an earlier study (Agelet et al., 2012). The instrument has also been found to be very sensitive to beans with cracked skin (results not shown). Preprocessing with SNV would diminish differences from physical characteristics between beans, but the previous research from Roussel et al. (2001) and unpublished results with the current data showed that this preprocessing method of the spectra decrease the ability to discriminate RR and conventional seeds. The FT-NIR, on the other side, may only be affected by the seed pathlength as only the central section of the bean is analysed; the diaphragm allows adjusting the seed in the middle of the detector. This leads to models which are not so specific to the seed morphological characteristics even without using preprocessing methods.

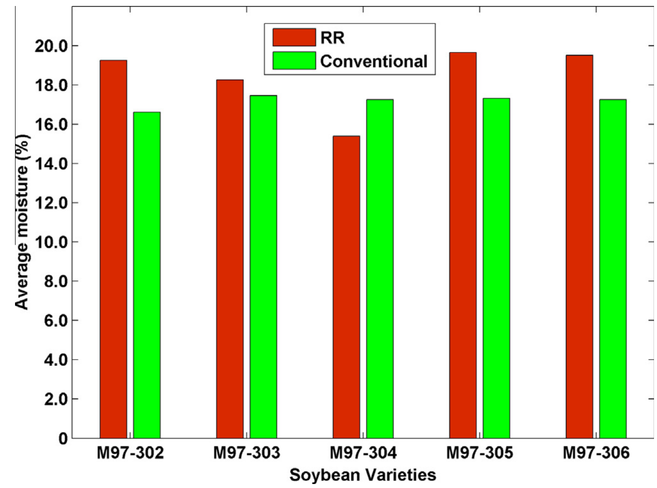

Fig. 1. Average moisture content of RR and conventional classes for each variety after increasing the moisture content by environmental absorption.

M97-303 and M97-304 varieties on the other hand, were very poorly discriminated by the other models and even by the model developed with beans from their same variety. Those varieties had in common their lower overall moisture content and lower moisture of RR class, compared with the other varieties. The model from M97-304 variety was also unable to discriminate any other variety, independently from the instrument utilised. Almost all conventional seeds from the other varieties are mainly classified as RR. Again, after finding overall differences in oil, protein and weight between RR and conventional seeds among all the varieties, the main difference of that variety when compared with the rest is its low overall moisture content and lower moisture content for the RR class. Accordingly, that variety has all the RR seeds classified as conventional when classified with models from the other varieties. The relevance of the wavelength where water absorption occurs seems to be stronger than the absorptions from oil or protein. This leads one to think that moisture-related interaction or other minor compounds and features happening at wavelengths where water absorption occurs, drives the discrimination between RR and conventional classes.

3.3. Moisture effect

Two PLS-DA models were developed, one for each instrument, including all the soybeans from all varieties except M97-304. For the light tube, the model had 8 latent variables and only 8 beans were incorrectly classified by cross-validation. When predicting beans from M97-304 variety, all RR beans were misclassified as conventional, and few conventional were misclassified as RR (Fig. 2). The FT-NIR model required 6 latent variables, and not all the varieties behave the same when developing the model. Half of the RR beans from the M97-303 variety could not be modeled as RR. Again, that variety is the one that after M97-304 has the lowest bulk predicted moisture for the RR counterpart. When the M97-304 beans were predicted, all conventional beans were correctly classified, and 89 of 147 RR beans were misclassified. The number of misclassified beans would probably have increased if variety M97-303 were removed from the training set.

Moisture content of the high moisture beans. After all the beans with higher moisture were scanned with both instruments, their true moisture was determined and recorded. The overall average moisture from both conventional and RR classes differed in almost one percentage point (17.2% for conventional and 18.4% for RR). And surprisingly, a similar pattern to what was observed with lower moisture beans was shown when examined by variety.

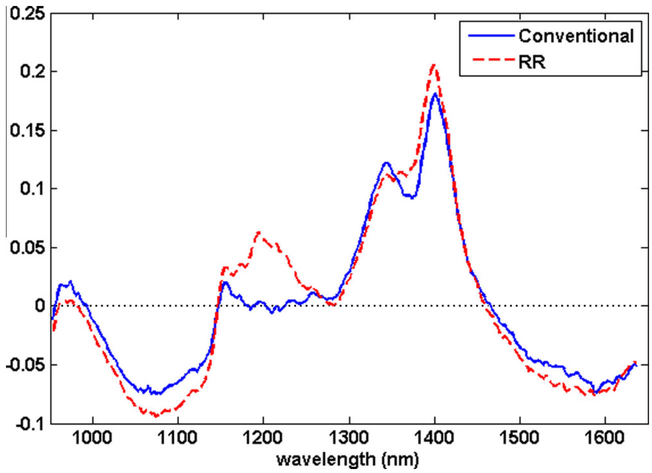

Fig. 2. Light tube subtraction spectra from conventional and RR beans showing moisture effect on beans. Subtraction spectra were obtained subtracting dry beans SNV averaged spectra from wet beans SNV averaged spectra.

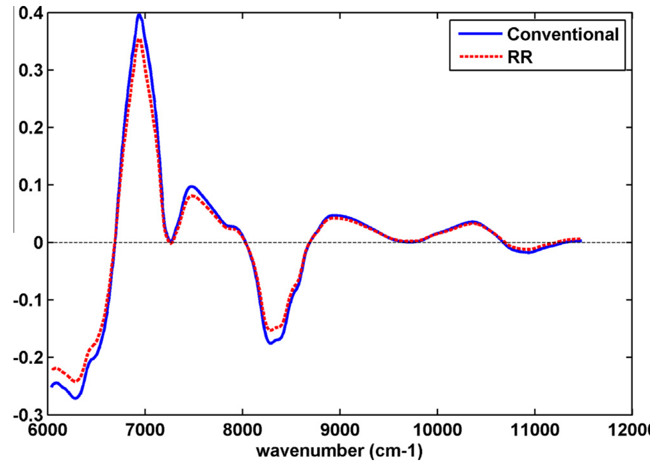

Fig. 3. FT-NIR subtraction spectra from conventional and RR beans showing moisture effect on beans. Subtraction spectra were obtained subtracting dry beans SNV averaged spectra from wet beans SNV averaged spectra.

Fig. 1 shows the average moisture for each variety and class (conventional, RR). Again, all varieties except M97-304 had higher moisture for the RR class. M97-302, M97-305, and M97-306 had over two percentage points of moisture difference between RR and conventional classes. M97-303, whose moisture for conventional and RR classes were closer for low moisture beans, had a 0.8 point difference between the RR and conventional classes. M97-304 again showed higher moisture for conventional beans (17.3%) compared to RR (15.4%).

These results confirm that for almost all varieties, RR and conventional beans behave different when exposed to high moisture conditions. RR beans seemed to absorb and retain more water than conventional. One variety in five used in the study showed the opposite behavior (variety M97-304). The explanation for this phenomenon may rely on a possible pleitropic effect of the RR gene on structural characteristics or concentration of minor compounds for certain soybean varieties. Research by some authors mentioned a possible pleiotropic effect of the RR gene on lignin production and thus differences in lignin content in soybean plants (Zobiole, Bonini, de Oliveira, Kremer, & Ferrarese-Filho, 2010; Gertz, Vencill, & Hill, 1999; Coghlan, 1999). Our study appears to agree with these authors in the fact that structural changes in soybean husk such as lignin or cellulose content would have an impact on bean hygroscopic properties.

Prediction of high moisture beans with the dry bean model. When high moisture beans ($n = 187$) from all varieties were predicted with the model developed with dry beans, all conventional beans were misclassified as RR and all RR beans were correctly classified. The results were the similar for both instruments. When focusing on beans from variety M97-304, the trend was the same. From the whole range, the conventional beans with the lowest moisture content from the group (8.5–10%) were closer to being correctly classified as conventional. However, and since their moisture content was slightly higher than the conventional beans in the training set, they were still misclassified as RR. These results again confirm the fact that moisture highly impacts the discrimination of conventional and RR beans. Even small differences in moisture content make a big difference, and somehow a bean with slightly higher moisture content is classified as RR bean. Considering that any compound such as oil, protein, and water is found in a somewhat wide range in a seed sample from the same variety and class, it would be incorrect to assume that the discrimination between conventional and RR beans is because of moisture alone. Overlap in moisture content between classes is expected, but still the classification is very sensitive to small moisture changes. In order to evaluate any possible interaction involving moisture, spectra were averaged from all the 'dry' beans for all varieties (around 1500 spectra) and processed the average spectrum with standard normal variate (SNV) to eliminate the offset and diminish other impacts such as light scattering from the spectrum. This was also done with the high moisture bean spectra (average + SNV). Averaged dry and wet spectra were then subtracted. In absence of interaction, only the peak from water absorption at the second overtone (around 1430 nm) should be noticeable. Fig. 2 shows the result of subtracting the dry SNV average spectrum from the wet SNV average spectrum, for both conventional and RR beans. There is a large water absorption peak at 1430 nm, but also a big peak around the wavelength 1350 nm. The band at 1365 nm has been assigned to stretching and deformation of CH bounds in cellulose (Fujimoto Matsumoto, Kurata, & Tsuchikawa, 2008). This could be explained by the fact that when increasing the moisture there are structural changes in the seed, and the seed fiber is stretched to cope with slight seed swelling. Another peak shows only for RR beans around 1150–1250 nm. This is also a second overtone stretch region for cellulose (Fujimoto Matsumoto, Kurata, & Tsuchikawa, 2008). Conventional and RR beans lead to slight different features when increasing moisture in the seed, all in the region of structural carbohydrates such as cellulose. The distinct peak of RR beans was not detectable in the FT-NIR transmittance spectra. Fig. 3 shows the full range transmittance spectra of the FT-NIR instrument, covering also the second overtone region. The largest peak (7000 cm^{-1}) corresponds to a water absorption region but also an absorption region for amorphous cellulose. An additional cellulose peak (7321 cm^{-1}) appears to be influenced by cellulose swelling and deformation. However, in our transmittance working range, the peak corresponding to water absorption (around 10,255 cm^{-1}) was quite broad and other peaks corresponding to carbohydrate changes may be overlapped and not as evident as for the spectra in reflectance mode.

4. Conclusions

In agreement with previous publications (McCann et al., 2005; Taylor et al., 1999) (Harrigan et al, 2010), no overall significant differences were found between RR and conventional soybean classes in composition of major compounds such as protein, oil, or seed weight. Within a single variety, however, the difference between RR and conventional beans may be significant based on our single crop-year study involving 5 varieties and 300 beans total (150 beans per class). The discrimination by Near Infrared within a single variety was accurate by PLS-DA models (>98% accuracy), but surprisingly the varieties that could be classified with highest accuracies were not the ones showing significant differences in oil, protein, or weight. Varieties that had the lowest bulk moisture predictions were the ones with more misclassifications. Although RR classes trend toward higher equilibrium moisture content than conventional classes, the differences were very small and a discrimination based only on moisture content is unlikely since the distribution of the concentration of any compound in a seed batch is wide enough to allow considerable overlap of classes. During seed conditioning to increase beans to higher moisture content, it was noticed that RR beans tend to absorb more water than their conventional counterparts and relative to the moisture characteristics of dry soybeans, RR classes also retain moisture longer with time. Among the relevant seed characteristics that contribute to permeability and water uptake are compounds which are part of the seed hull, mainly carbohydrates (cellulose, hemicellulose, lignin). This would suggest that for most of the varieties there exists a pleitropic effect of the introduced RR gene which impacts the seed hull or structural characteristics. We have seen how water, besides having an overlapping NIR absorption band with carbohydrate regions, causes physical changes in the seed hull do to cellulose deformation. Those phenomena interfere with the differences in the hull composition or structure that may exist between RR and conventional beans, and for this reason any small change in bean moisture has a big impact in the discrimination.

It was demonstrated that NIRS can detect small differences between RR and conventional soybeans. These differences may be due to small compositional changes in carbohydrates or structural differences in the seed hull. Future experiments could involve microscopic examination of the seed coats to find out more about these possible differences in hull structure.

Moisture induces physical changes within the hull as well as occupying a NIR spectral region that overlaps with carbohydrates. The complexity of this interaction makes the use of NIRS for discriminating conventional and RR beans not suitable for traders or in-field measurements, and only reliable for measurements at small and well-controlled moisture ranges.

Acknowledgments

The author acknowledges the helpfulness and problem solving efficiency through this research of Buchi personnel Gael Rolland, Bradley C. Miller, and Rolland L. Rubinovitz.

References

Agelet, L. E., Gowen, A. A., Hurburgh, C. R., & O'Donell, C. (2012). Feasibility of conventional and Roundup Ready® soybeans discrimination by different near infrared reflectance technologies. *Food Chemistry, 134*(2), 1165–1172.

Armstrong, P. R. (2006). Rapid single-kernel NIR measurement of grain and oil-seed attributes. *Applied Engineering in Agriculture, 22*, 767–772.

Bakshi, A. (2003). Potential adverse health effects of genetically modified crops. *Journal of Toxicology and Environmental Health, Part B, 6*(3), 211–225.

Benbrook, C. (2009). *Impacts of genetically engineered crops on pesticide use: The first thirteen years*. <http://www.organic-center.org/science.pest.php?action=view&report_id=159> (accessed 08.11.12).

Coghlan, A. (1999). Splitting Headache. New Scientist, 20. Nov. 1999.

Cohen, S. N., Chang, A. C. Y., Boyer, H. W., & Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences, 70*, 3240–3244.

Elmore, R. W., Roeth, F. W., Nelson, L. A., Shapiro, C. A., Klein, R. N., & Martin, A. (tul=0?>Elmore et al 2001). Glyphosate-resistant soybean cultivar yields compared with sister lines. *Agronomy Journal, 93*, 408–412.

Fujimoto, T., Kurata, Y., Matsumoto, K., & Tsuchikawa, S. (tul=0?>Fujimoto et al 2008). Application of near infrared spectroscopy for estimating wood mechanical properties of small clear and full length lumber specimens. *Journal of Near Infrared Spectroscopy, 16*, 529–537.

Gertz, J.M., Vencill, W.K., & Hill, N.S. (1999). Tolerance of Transgenic Soybean (Glycine max) to Heat Stress. British Crop Protection Conference – Weeds, 15–19 Nov 1999. Brighton: 835–840.

Harrigan, G. G., Lundry, D., Drury, S., Berman, K., Riordan, S. G., Nemeth, M. A., et al. (tul=0?>Harrigan et al 2010). Natural variation in crop composition and the impact of transgenesis. *Nature Biotechnology, 28*, 402–404.

Hübner, P., Waiblinger, H.-U., Pietsch, K., & Brodmann, P. (2001). Validation of PCR methods for the quantification of genetically modified plants in food. *Journal of the Association of Official Analytical Chemists International, 84*, 1855–1864.

Jimenez, J. J., Bernal, J. L., Nozal, M. J., Toribio, L., & Bernal, J. (2009). Profile and relative concentrations of fatty acids in corn and soybean seeds from transgenic and isogenic crops. *Jounal of Chromatography A, 1216*, 7288–7295.

Konduru, S., Kruse, J., & Kalaitzandonakes, N. (2008). The global economic impacts of Roundup Ready soybeans. In G. Staycey (Ed.). *Plant genetic/genomics, genetics and genomics of soybeans* (Vol. 2 (Part IV), pp. 375–395). New York: Springer.

Lee, J. H., & Choung, M. G. (2011). Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. *Food Chemistry, 126*, 368–373.

McCann, M. C., Liu, K., Trujillo, W. A., & Dobert, R. C. (2005). Glyphosate-tolerant soybeans remain compositionally equivalent to conventional soybeans (*Glycine max* L.) during three years of field testing. *Journal of Agriculture and Food Chemistry, 53*(13), 5331–5335.

Miljuš-Djukić, J., Banović, B., Jovanović, Z., Majić, D., Milisavljević, M., Samardžić, J., et al. (2010). Abundance of soybean Roundup Ready modification in food and feed samples from Serbian retail markets. *Romanian Biological Letters, 15*(1), 102–109.

Roussel, S. A., Hardy, C. L., Hurburgh, C. R., Jr., & Rippke, G. R. (2001). Detection of Roundup Ready™ soybeans by near-infrared spectroscopy. *Applied Spectroscopy, 55*(10), 1425–1430.

Schenepf, R. (2003). *Genetically engineered soybeans: Acceptance and intellectual property rights, issues in South America*. <http://www.cnie.org/NLE/CRSreports/03Oct/RS21558.pdf> (accessed 08.11.12).

Tallada, J. G., Palacios-Rojas, N., & Armstrong, P. R. (2009). Prediction of maize seed attributes using a rapid single kernel near infrared instrument. *Journal of Cereal Science, 50*, 381–387.

Taylor, N. B., Fuchs, R. L., MacDonald, J., Shariff, A. R., & Padgette, S. R. (1999). Compositional analysis of glyphosate-tolerant soybeans treated with glyphosate. *Journal of Agricultural and Food Chemistry, 47*, 4469–4473.

Underwood, A. J. (1997). *Experiments in ecology: Their logical design and interpretation using analysis of variance*. Cambridge: Cambridge University Press.

Zobiole, L. H. S., Bonini, E. A., de Oliveira, R. S., Kremer, R. J., & Ferrarese-Filho, O. (2010). Glyphosate affects lignin content and amino acid production in glyphosate-resistant soybean. *Acta Physiologiae Plantarum, 32*, 831–837.